Argonne NATIONAL LABORATORY

... for a brighter future

ANL Tier3 & ASC computer infrastructure

S.Chekanov

(HEP Division, ANL)

DOE review, Sep. 2009

U.S. Department of Energy

UChicago Argonne LLC

A U.S. Department of Energy laboratory managed by UChicago Argonne, LLC

ANL Tier3 & ASC. Present

http://atlaswww.hep.anl.gov/

~90 registered users

Hardware:

- **50 cores for all users**
- **Includes 2 cluster prototypes:**
  - 24 cores/6TB ArCond (developed by ANL)
    - *Currently used for analysis*
  - 12 cores (includes Duke's computers)
    - *SRM/Xrootd (under development)*

Maintained by the HEP computing group:

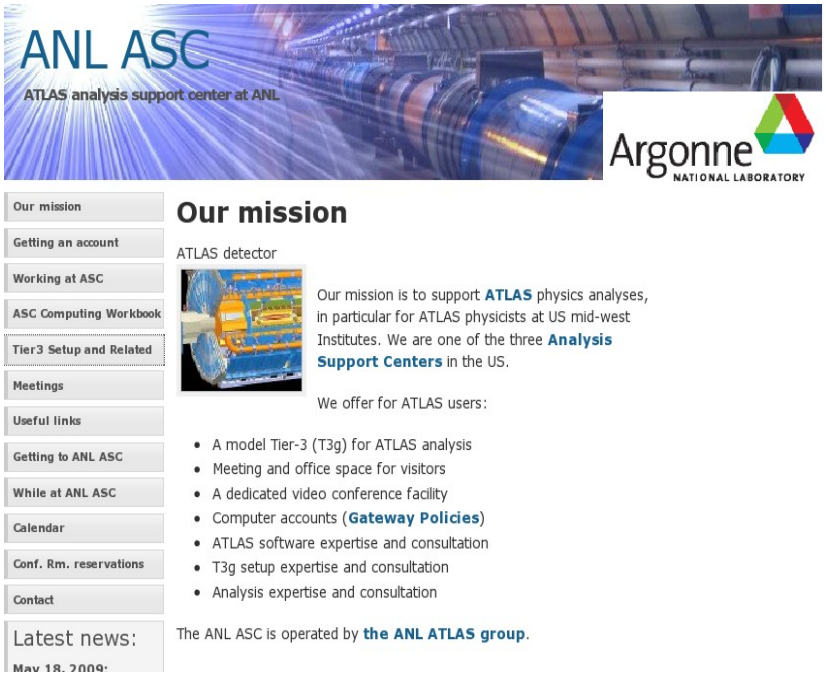
S.C* (leader)

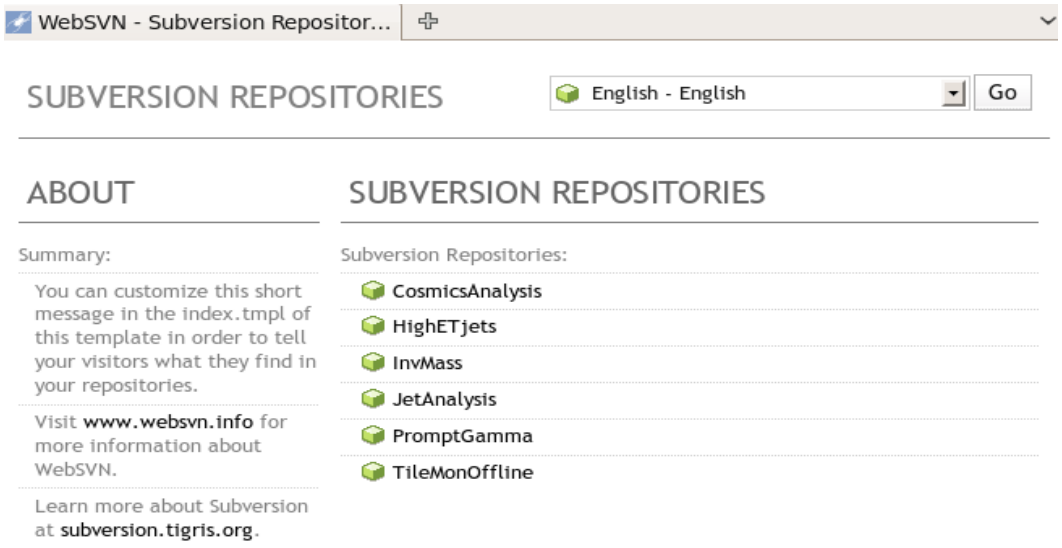
B.Blair*,E.Kovacs, J.Hinthorn,T.Hayden*

(* part time)

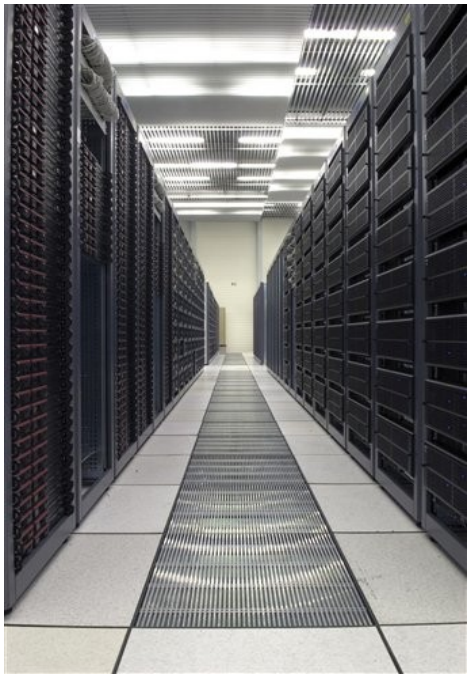
ASC software infrastructure:

- Atlas software releases
- Grid access
- Indico, Forum, Twiki with:
  - *ASC computer workbook*
  - *Tier3 setup guide*
  - *CVS, SVN, SVN browsers*
- 6 supported packages





Requirements for Tier3 cluster (T3g)



- Interactive & chaotic analyses
- No resource allocation and file staging for each job execution
  - faster data processing compared to the grid
  - reliability
- Low cost: tens of $k
- Off-the-shelf hardware
- Small effort in management (0.2FTE)
- No special network requirement & computer room
- Scalable, no I/O bottleneck

- **Prototype was designed (24 cores) based on the Condor and operational since Sep. 2008**
  - Fully satisfies the above requirements
- **7000 jobs since Sep. 2008**
  - Current fault rate <0.05%
- **5 ATLAS notes using the farm**

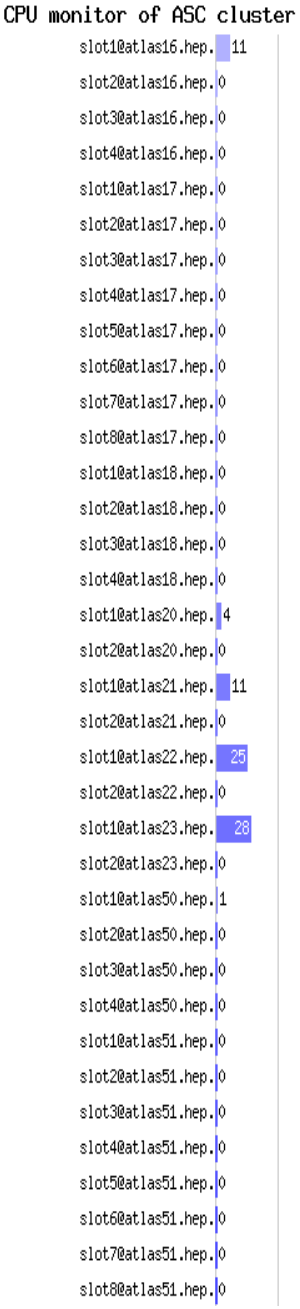
ANL PC farm (PCF) design

- **Easy to use**
  - **All you need to know is:**
    - *Which data set is available?*
    - *What software package to run?*
  - **Don't need to know:**
    - *Where the data are?*
    - *Where your jobs will run?*

Software package: ArCond – "Argonne**+Cond**or" **for Tier3s computer farms:**

- Python front-end of Condor for: job submission, data discovery, results retrieval
- Developed and supported at ANL ASC
- Will be a part of USATLAS Tier3 toolkit
- Documentation: http://atlaswww.hep.anl.gov/asc/arcond/

- **ANL PCF design will be a key for many US ATLAS Tier3 sites**
  - Can be in combination with xrootd file system
    - under development → help from D.Benjamin



Data sets on the PC farm

- **Since Sep. 2008, we store 17422 AOD MC files**
  - ~ 4M Monte Carlo AOD events (+ few ESD sets)
  - Corresponds to ~**25%** of the total capacity of the PC farm prototype

/data1/mc/gamma_jet/pt17/AOD	atlas52	gamma+jet samples, r14.2, pt>17 GeV. Also available: pt40, pt8 pt600
/data1/mc/pythia_gfilter/pt17/AOD	atlas51	Filtered background sample, r14.2, pt>17 GeV. Also available: pt pt400, pt600
/data1/mc/PythiaZeegam25/AOD	atlas51-52	Z+gamma+X samples, r14.2, pt>25 GeV
/data1/mc/BaurZeegam/AOD	atlas51	Z+gamma+X, Baur MC, r14.2, pt>25 GeV, X-section=463.622 p each file
/data1/mc/mc08.105802.JF17_pythia_jet_filter.recon.AOD.e347_s462_r541/AOD	atlas51-53	~1.5 M events, inc.Pythia after JetFilter, r14.2, pt>17
/data1/mc/mc08.106070.PythiaZeeJet_Ptcut.recon.AOD.e352_s462_r541/AOD	atlas51-53	Z->e+e- + jet events, r14.2.20, 250 events in each file, 797 files, 968.637 pb, efficiency = 0.90
/data1/mc/mc08.106071.PythiaZmumuJet_Ptcut.recon.AOD.e352_s462_r541/AOD	atlas51-53	Z->mu+mu- + jet events, r14.2.20, 250 events in each file, 791 file 968.637 pb, efficiency = 0.90
/data1/mc/mc08.106072.PythiaZtautauJet_Ptcut.recon.AOD.e352_s462_r541/AOD	atlas51-53	Z->tau+tau- + jet events, r14.2.20, 250 events in each file, 759 file 968.637 pb, efficiency = 0.90
/data1 /mc/mc08.106379.PythiaPhotonJet_AsymJetFilter.recon.AOD.e347_s462_r541/AOD	atlas51-53	250k events, gamma+jet, ckin(3)>15 GeV
/data1/mc/MC08/JS0/ESD	atlas53	also JS1, JS2,JS3,JS4,JS5,JS6,JS7 available. Talk to Belen a
/data1/mc/mc08.107141.singlepart_pi0_Et40.recon.AOD.e342_s439_r546/AOD	atlas51	200 files, r14.2.20.3, single pi0
/data1/mc/mc08.107041.singlepart_gamma_Et40.recon.AOD.e342_s439_r546/AOD	atlas51	189 files, r14.2.20.3, single gamma
/data1/mc/mc08.107680.AlpgenJimmyWenuNp0_pt20.recon.AOD.e349_a68/AOD	atlas51-53	1202 files, r14.2.20, W->e+nu+0 partons
/data1/mc/mc08.107681.AlpgenJimmyWenuNp1_pt20.recon.AOD.e349_a68/AOD	atlas51	242 files, r14.2.20, W->e+nu+1 partons
/data1/mc/mc08.107682.AlpgenJimmyWenuNp2_pt20.recon.AOD.e349_a68/AOD	atlas51	624 files, r14.2.20, W->e+nu+2 partons
/data1/mc/mc08.107683.AlpgenJimmyWenuNp3_pt20.recon.AOD.e349_a68/AOD	atlas51	165 files, r14.2.20, W->e+nu+3 partons
/data1/mc/mc08.107684.AlpgenJimmyWenuNp4_pt20.recon.AOD.e349_a68/AOD	atlas51	48 files, r14.2.20, W->e+nu+4 partons
/data1/mc/mc08.107685.AlpgenJimmyWenuNp5_pt20.recon.AOD.e349_a68/AOD	atlas51	22 files, r14.2.20, W->e+nu+5 partons

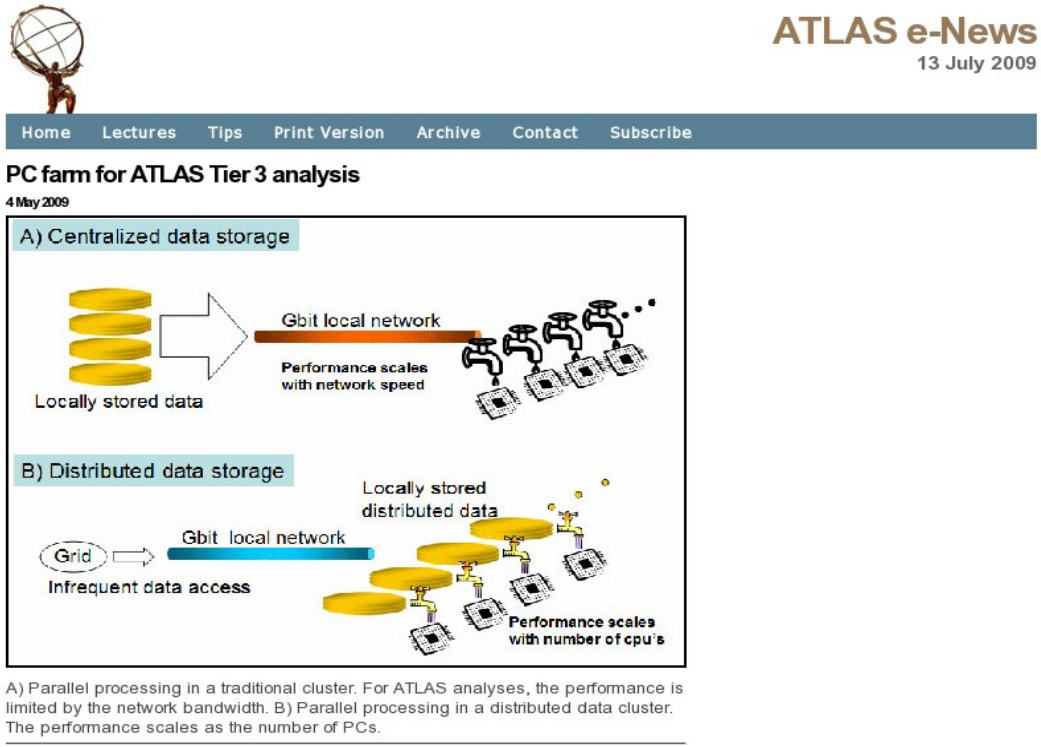
FDR2 reprocessed data: ||

/data1/mc/fdr08_run2.0052280.physics_Egamma.recon.AOD.o3_f47_r575/AOD	atlas51-53	FDR2 AOD data, release 14.2.24
/data1/mc/fdr08_run2.0052280.physics_Egamma.recon.DPD_CALOJET.o3_f47_r575/AOD	atlas51-53	FDR2 DPD data, release 14.2.24
/data1/mc/fdr08_run2.0052280.physics_Egamma.recon.DPD_EGAMMA.o3_f47_r575/AOD	atlas51-53	FDR2 DPD data, release 14.2.24
/data1/mc/fdr08_run2.0052280.physics_Egamma.recon.DPD_PHOTONJET.o3_f47_r575/AOD	atlas51-53	FDR2 DPD data, release 14.2.24
/data1/mc/fdr08_run2.0052280.physics_Jet.recon.AOD.o3_f47_r575/AOD	atlas51-53	FDR2 AOD data, release 14.2.24

PC farm challenge for T3g sites

- **A complete T3G PC farm setup is given on the ANL ASC page (atlaswww.hep.anl.gov):**

Article in ATLAS e-News



More details: "A PC farm for ATLAS Tier3 analysis"
S.C., R.Yoshida, ATL-COM-GEN-2009-016

Getting data from Tier1/2 to ASC ANL

T2 Site	Tuning 0	Tuning 1
AGLT2_GROUPDISK	-	62 Mbps log
BNL-OSG_GROUPDISK	52 Mbps log	272 Mbps log
SLACXRD_GROUPDISK	27 Mbps log	347 Mbps log
SWT2_CPG_GROUPDISK	36 Mbps log	176 Mbps log
NET2_GROUPDISK	83 Mbps log	313 Mbps log
MWT2_UC_MCDISK	379 Mbps log	423 Mbps log

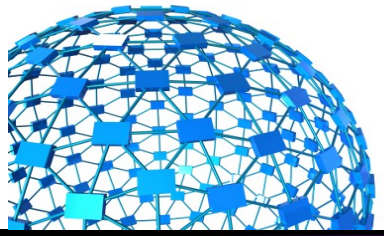
SL 5.3 TCP tune Recommended by ESnet

Brown color: at least one file has 0 size

Benchmarks by R.Yoshida & D.Benjamin for Duke's T3 site

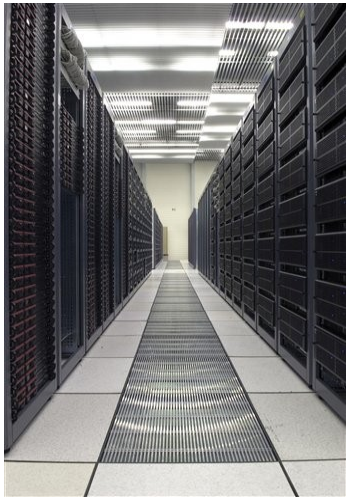
Satisfactory download rates for several Tier1-2 sites: 3-4 TB/day (after the ESnet tune)

All tests and TCP tunes are documented and available for Tier3 sites

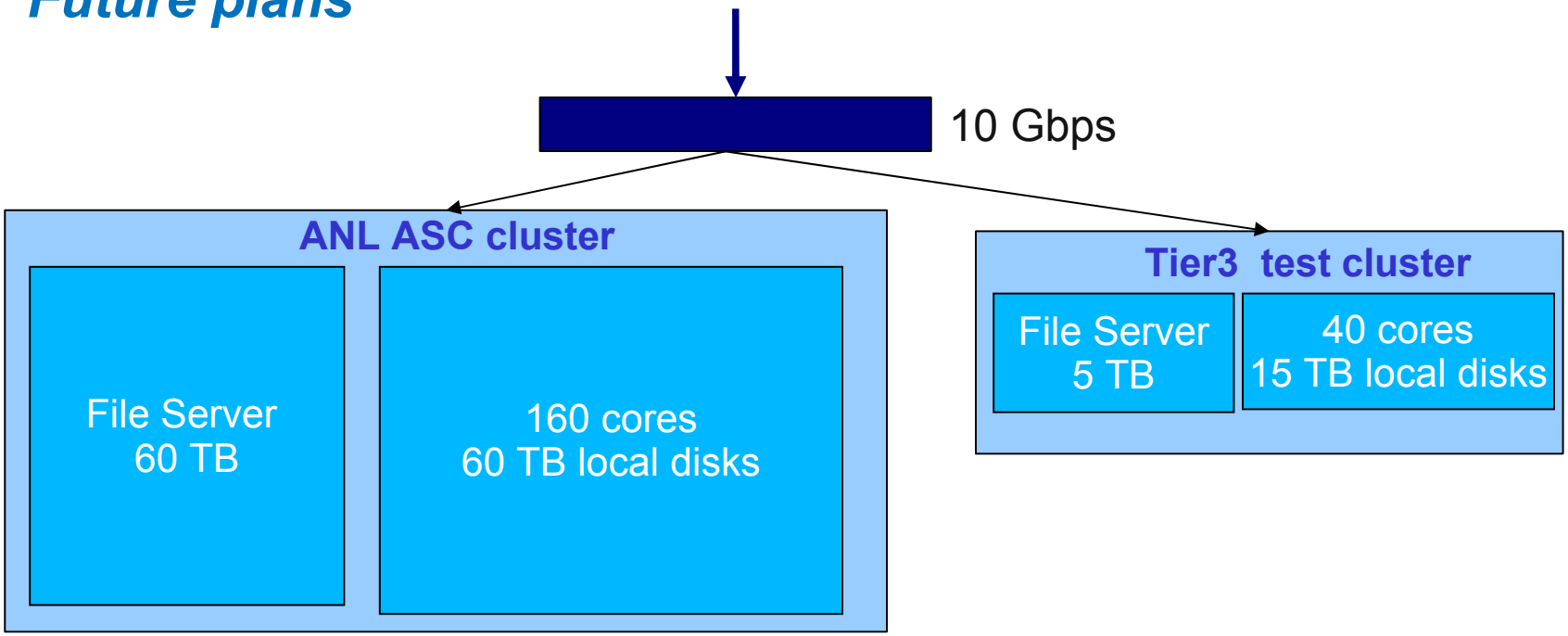


Future plans

- **Expand current capabilities of ANL ASC computing**
  - current infrastructure is not adequate when data will arrive
- **Increase the size and flexibility of the existing test T3g cluster**
- **Further develop the model for Tier3 computing for US ATLAS**
- **Help to support other Tier3 sites**



Future plans



- **factor ten increase in the number of processing cores**
- **ANL ASC design for both clusters (ArCond/Condor)**
  - use xrootd file system for the test cluster (under development)
- **file servers for data storage**
- **consolidate both clusters into a single 200 core cluster**
- **10 Gb connection**
- **build a computer room**

Backup slides:

Expected performance:
24 Xeon 5404 cores (now) vs 200 Xeon 5520 (future)

5500 (Nehalem) processors are 50%-100% faster than Harpertown Xeon (5400)
Assume 50% (benchmarks are coming):

- **Running over AOD files**
  - 0.5M events /h → **6M/h**
- **Fast MC simulation and on the fly analysis**
  - 1.5M events /h → **18M/h**
- **Running over C++/ROOT ntuples**
  - 1000M events /h (1M events / min for 1 core). **10B? I/O limit?**
- **Generating MC truth ntuples**
  - 2.5M events /h → **30M/h**
- **AOD production (generating & reconstructing MC events)**
  - 120 events /h → **1400/h**